

# My Report

This is a simulation exercise to compare sample mean with theoretical mean of the distribution. In order to investigate if repeated samples from exponential distribution match the expected outcomes , a simulation of 1000 random samples is used, from an exponential distribution with lambda = 0.2. The sample means will be compared to theoretical mean to see if they overlap.

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.2

library(knitr)

## Warning: package 'knitr' was built under R version 3.6.2

lambda <- 0.2
n <- 40
num_sim <- 1000

data <- matrix(rexp(n=num_sim*n, rate=lambda),num_sim, n)

sample_mean <- rowMeans(data)

actual_mean<- mean(sample_mean)

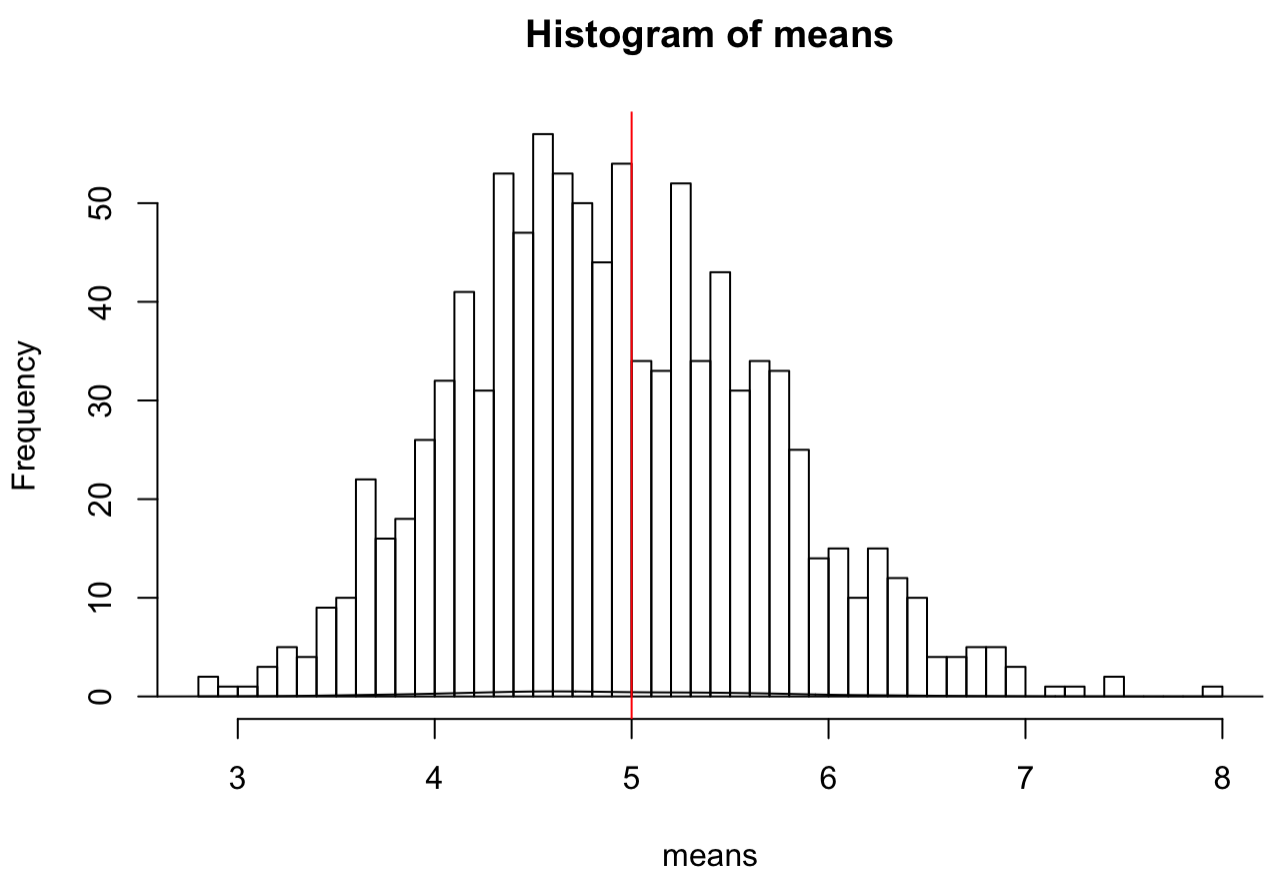
theoretical_mean <- 1/lambda
result <- data.frame("Mean"=c(actual_mean, theoretical_mean),
                     row.names = c("Sample Mean", "Theoretical mean"))
result

##           Mean
## Sample Mean   4.987162
## Theoretical mean 5.000000
```

The sample mean is very close to the theoretical mean of 5.  
Below we can see a histogram plot of an exponential distribution with n = 1000

```
means <- vector()

for (i in 1:1000){
  means <- c(means, mean(rexp(n, lambda)))
}
hist(means, breaks=40)
lines(density(means))
abline(v=1/lambda, col="red") # The red line indicates the mean
```



The theoretical variance equals 0.625. Sample variance means are close to overlap with the expected theoretical variance.

```
variance <- var(sample_mean)

theoretical_variance <- (1/lambda)^2 / n

outcome <- data.frame("Variance" = c(variance, theoretical_variance),
                     row.names = c("Sample variance", " Theoretical variance"))

outcome

##           Variance
## Sample variance   0.5946612
## Theoretical variance 0.6250000
```

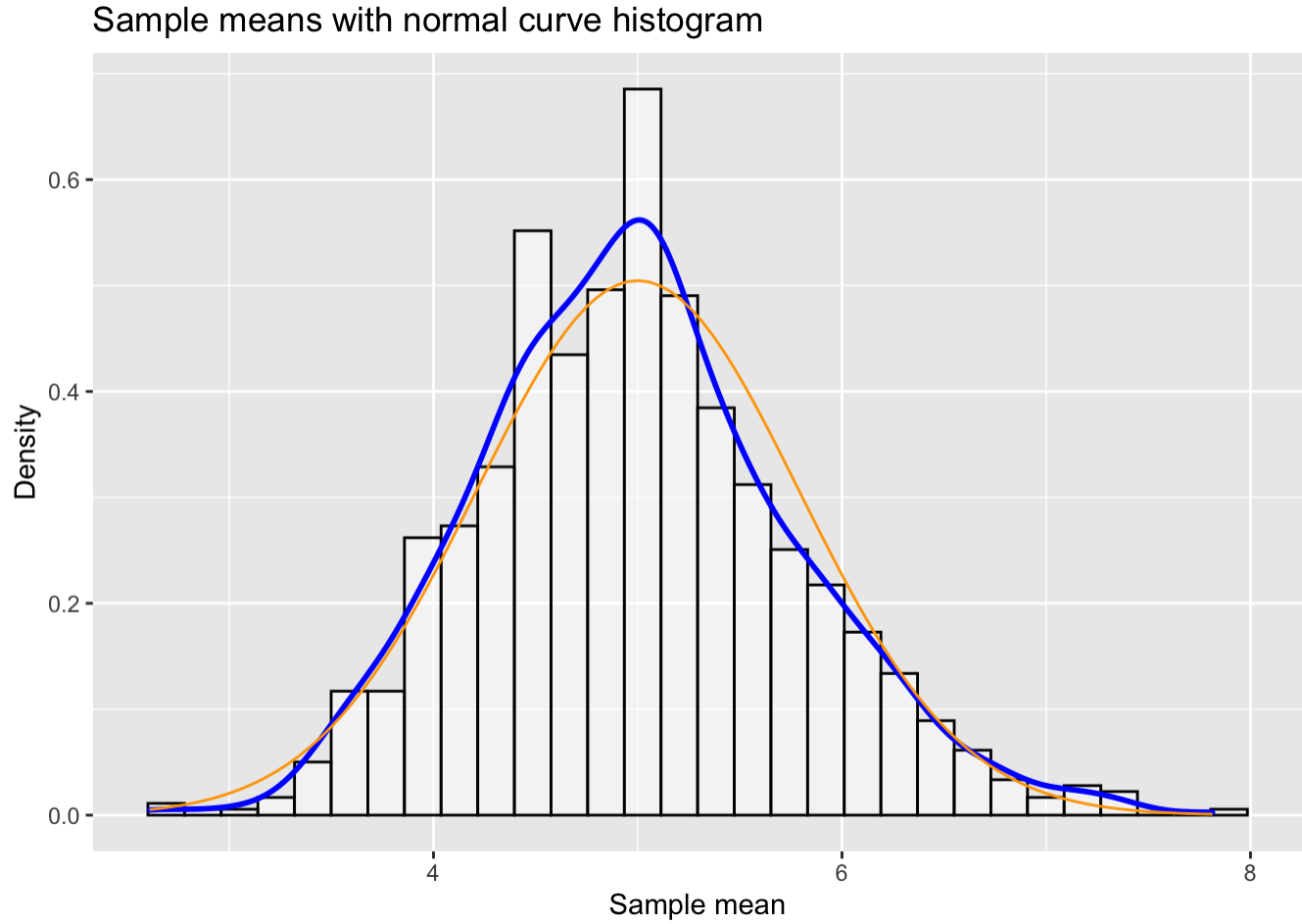
## Distribution

According to the Central Limit Theorem, the sample averages follow normal distribution. The following plot confirms normal distribution is almost matched by the sample means distribution.

```
sample_mean_data <- as.data.frame(sample_mean)

ggplot(sample_mean_data, aes(sample_mean)) +
  geom_histogram(aes(y=..density..), alpha=.5, position="identity", fill="white", col="black")+ geom_density(colour="blue", size=1)+
  stat_function(fun= dnorm, colour="orange", args=
    list(mean=theoretical_mean, sd=sqrt(theoretical_variance))) +
  ggtitle("Sample means with normal curve histogram")+
  xlab("Sample mean")+
  ylab("Density")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Normal Probability Plot of Residuals confirms distribution of sample means almost matches the theoretical normal distribution.

```
qqnorm(means)
qqline(means)
```

